The effect of individual differences in factor loadings on the standard factor model

Kelderman, H.; Molenaar, P.C.M.

published in
Multivariate Behavioral Research
2007

DOI (link to publisher)
10.1080/00273170701382997

document version
Publisher's PDF, also known as Version of record

MULTIVARIATE BEHAVIORAL RESEARCH, *42*(3), 435–456

The Effect of Individual Differences in Factor Loadings on the Standard Factor Model

Henk Kelderman
Free University, Amsterdam

Peter C. M. Molenaar
Pennsylvania State University, Pennsylvania

It is shown that the population-covariance matrix of a heterogeneous factor model may be indistinguishable from that of a standard factor model and that the standard likelihood-ratio goodness-of-fit statistic has but little power in detecting loading heterogeneity. The relation between loading heterogeneity and factor score reliability is studied and it is recommended that non-normality of the test-score distributions be tested to use factor scores with more confidence. Substantive justifications for the model assumptions and model-based methods to test specific hypotheses about the loading distribution, are discussed.

In applied psychology, factor analysis is often used to develop diagnostic instruments. To obtain a good measure of a construct of interest (Cronbach & Meehl, 1955), a calibration study is performed wherein a battery of tests is administered to a large sample of subjects. Under the standard assumptions of multivariate normality of factors and residuals, test statistics and parameter estimates are computed from the covariance matrix (Jöreskog, 1971; Lawley & Maxwell, 1971). If the model fits the data, the parameter estimates are used to compute new subjects' factor scores and their confidence intervals to determine their position on the construct (Mellenbergh, 1994, 1996).

In this paper, it is shown that a well-fitting factor model thus obtained does not necessarily mean that the essential assumptions hold and that factor scores

The authors would like to thank an anonymous reviewer, and the editor, Roger Millsap, for their helpful comments on previous drafts and Conor V. Dolan for help with some of the statistics and graphics.

Correspondence concerning this article should be addressed to Henk Kelderman, 2E33 Transitorium, Vrije Universiteit, van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. E-mail: h.kelderman@psy.vu.nl

and their standard errors may safely be used for diagnostic purposes. It is shown that heterogeneity of factor loadings is a serious threat to the usefulness of the factor model, which may not be detected in the standard calibration procedure. Some implications and solutions are discussed.

In what follows, we introduce the standard factor model (SFM) and its heterogeneous variant (HFM). It is shown analytically that the population-covariance matrices from both models are identical if the loadings of the HFM are normally distributed and uncorrelated. In addition, simulated sample-covariance matrices are shown to yield almost indistinguishable likelihood-ratio fit statistics. In another simulation study, it is shown to what extend the reliability of factor scores deteriorates as a result of loading heterogeneity. To use factor scores with more confidence, it is proposed that non-normality of the test-score distributions be tested first. The performance of Shapiro Wilk's W test for non-normality is discussed and its relation to the reliability of factor scores studied. W tests against a general alternative and does not assume a specific distribution for the loadings. We discuss model-based methods to test specific hypotheses about the loading distribution, and substantive justifications for the model assumptions that we make in this paper.

THE FACTOR MODEL

Denote by y_{ij} the scores on n measurements $j = 1, \ldots, n$, of a randomly selected individual i from population Π, by η_i the individual's value on a common factor, by λ_j the loading of measurement j on the factor, and by ϵ_{ij} the residual. For simplicity, we assume that measurements $\{y_{ij}\}_j$ each have mean zero. In the standard 1-factor model (S1FM, Spearman, 1904), it is assumed that, in Π,

$$y_{ij} = \lambda_j \eta_i + \epsilon_{ij}, \;\; j = 1, \ldots, n, \tag{1}$$

and

$$\operatorname{cov}(\epsilon_{ij}, x_i) = 0, \;\; j = 1, \ldots, n, \tag{2}$$

where x_i stands for η_i or $\epsilon_{ij'}$ ($j' = 1, \ldots, j-1, j+1, \ldots, n$). Furthermore, without loss of generality, assume throughout this paper that

$$\mathrm{E}(\eta_i) = 0, \;\; \operatorname{var}(\eta_i) = 1, \;\text{ and } \mathrm{E}(\epsilon_{ij}) = 0, \;\; j = 1, \ldots, n. \tag{3}$$

The S1FM is easily generalized to a model with multiple factors by writing it in matrix algebra (Lawley & Maxwell, 1971).

In (1), the loadings are assumed fixed and the same for all individuals in Π. However, in some applications, such as quantitative behavioral genetics, there are pervasive substantive arguments in favor of individual differences in factor loadings (Molenaar, Huizenga, and Nesselroade, 2000).

Now suppose that the true structure is a heterogeneous 1-factor model (H1FM):

$$y_{ij} = \lambda_{ij}\eta_i + \psi_{ij}, \; j = 1, \ldots, n, \tag{4}$$

where λ_{ij} is the loading of a randomly selected individual from Π. It is assumed that

$$\text{cov}(\psi_{ij}, x_i) = 0, \; j = 1, \ldots, n, \tag{5}$$

where x_i stands for η_i, λ_{ij} $(j = 1, \ldots, n)$ and $\psi_{ij'}$ $(j' = 1, \ldots, j-1, j+1, \ldots, n)$. For simplicity, but without loss of generality, let the true model be normalized such that

$$\text{E}(\eta_i) = 0, \; \text{var}(\eta_i) = 1, \; \text{and } \text{E}(\psi_{ij}) = 0, \; j = 1, \ldots, n. \tag{6}$$

If maximum likelihood estimates and likelihood-ratio testing are desired, the usual distributional assumption of the standard model (1) is that the factor η_i and residuals $\{\epsilon_{ij}\}_j$ have a joint multivariate normal distribution subject to restriction (2). Model (4) is a random-loadings factor model. If it is assumed that the factor loadings are completely independent and normally distributed, we have the random-loadings factor model described by Ansari, Jedidi and Dube (2002). In the next section we show that under the weaker assumption that normally distributed loadings are independent of *each other*, the population-covariance matrix of a HFM is the same as that of a SFM.

POPULATION COVARIANCE STRUCTURES OF HETEROGENEOUS AND STANDARD FACTOR MODELS MAY BE IDENTICAL

Let G_1 denote the set of S1F structures that satisfy (1) and G_2 both (1) and (2). Similarly, H_1 satisfies (4) and H_2 satisfies (4) and (5). For simplicity, let $\text{D}(x_i) \equiv x_i - \text{E}(x_i)$ denote the deviation score for some random variable x_i. The next lemma shows that heterogeneous loadings can be absorbed by the S1FM.

Lemma 1. *The set of heterogeneous 1-factor structures $E_1 \subset H_1$ whose elements satisfy*

i) $\lambda_j = \text{E}(\lambda_{ij})$,
ii) $\epsilon_{ij} = \text{D}\{\text{D}(\lambda_{ij})\eta_i\} + \psi_{ij}$,

is equivalent in Π with G_1.

Proof. Using (i), (ii), and $x_i = \mathrm{E}(x_i) + \mathrm{D}(x_i)$, one has

$$\begin{aligned}\lambda_{ij}\eta_i + \psi_{ij} &= \mathrm{E}(\lambda_{ij})\eta_i + \mathrm{D}(\lambda_{ij})\eta_i + \psi_{ij}\\ &= \mathrm{E}\{\mathrm{E}(\lambda_{ij})\eta_i + \mathrm{D}(\lambda_{ij})\eta_i\} + \mathrm{D}\{\mathrm{E}(\lambda_{ij})\eta_i + \mathrm{D}(\lambda_{ij})\eta_i\} + \psi_{ij}\\ &= \mathrm{E}\{\mathrm{E}(\lambda_{ij})\eta_i\} + \mathrm{E}\{\mathrm{D}(\lambda_{ij})\eta_i\} + \mathrm{D}\{\mathrm{E}(\lambda_{ij})\eta_i\} + \epsilon_{ij}\\ &= \mathrm{E}(\lambda_{ij})\mathrm{E}(\eta_i) + \mathrm{cov}(\lambda_{ij}, \eta_i) + \mathrm{E}(\lambda_{ij})\mathrm{D}(\eta_i) + \epsilon_{ij} \qquad (7)\\ &= \lambda_j\mathrm{E}(\eta_i) + \mathrm{cov}(\lambda_{ij}, \eta_i) + \lambda_j\mathrm{D}(\eta_i) + \epsilon_{ij}\\ &= \mathrm{cov}(\lambda_{ij}, \eta_i) + \lambda_j\eta_i + \epsilon_{ij}\\ &= \lambda_j\eta_i + \epsilon_{ij}\end{aligned}$$

for $i \in \Pi$ and $j = 1, \ldots, n$. Because, without loss of generality $\mathrm{E}(y_{ij}) = 0$, $\mathrm{E}(\eta_i) = 0$, and $\mathrm{E}(\epsilon_{ij}) = 0$, and $\mathrm{cov}(\lambda_{ij}, \eta_i)$ can be absorbed in the mean of y_{ij}. This proves the equivalence of E_1 and G_1 under the assumptions (i)–(ii).

To prove both models' equivalence of the covariance matrices, let $\sigma_{jk} = \mathrm{cov}(y_{ij}, y_{ik})$ be the covariance between measurement j and k under a certain factor model in a certain population Π, and let $\mathbf{\Sigma} = \{\{\sigma_{jk}\}_j\}_k$ be the full population (co)variance matrix. Furthermore, let $\mathcal{S}$ be the covariance structure of a certain factor model, that is, the set $\{\mathbf{\Sigma}\}$ of all possible covariance matrices that can be produced from that factor model in all possible populations. In the next theorem we state conditions under which the covariance structure of a the H1FM is the same as that of the S1FM.

Theorem 1. *The set $\mathcal{S}^{(E_2)}$ of H1FMs, $E_2 \subset H_2$, for which*

i) its random parameters have a multivariate normal distribution,
ii) $\mathrm{cov}(\lambda_{ij}, \lambda_{ik}) = 0$,

for $j \neq k = 1, \ldots, n$, is identical to the covariance structure $\mathcal{S}^{(G_2)}$ for the S1FMs G_2.

The proof of this theorem is in the Appendix I.

The results can be generalized to multiple factor analysis. In the proof of Lemma 1, λ_j, λ_{ij}, and η_i can be replaced by the row vectors $\boldsymbol{\lambda}_j^T$ and $\boldsymbol{\lambda}_{ij}^T$ and the column vector $\boldsymbol{\eta}_i$ respectively. First, write the residuals as uncorrelated factors with variances equal to one and with loadings equal to the residual standard deviations. Let F be the set of $|F| \geq n+1$ factor numbers and $r \in F$. Let $\mathbf{\Sigma}^{(F)}$ and $\mathbf{\Sigma}^{(r)}$ be the covariance matrices of a multiple-factor model and a single-factor model respectively. If all factors are uncorrelated it is easily seen that the multiple factor model can be written recursively as

$$\boldsymbol{\Sigma}^{(F)} = \boldsymbol{\Sigma}^{(F-r)} + \boldsymbol{\Sigma}^{(r)}$$

Therefore, by induction, if Theorem 1 holds for the single factor case, it also holds for the orthogonal multiple factor case. Since the oblique multiple factor model can be obtained by a one to one-to-one transformation of the orthogonal case, Theorem 1 also holds for the general case. Thus, covariance matrices from HFMs are identical to a corresponding SFM if its loadings are normally distributed and uncorrelated.[1]

This result, of course, does not necessarily hold for sample covariance matrices. In the next section we discuss the sampling properties and show that simulated sample-covariance matrices of both models yield almost indistinguishable likelihood-ratio fit statistics.

THE EFFECT OF LOADING HETEROGENEITY ON STANDARD GOODNESS OF FIT TESTING

As the sample grows larger, the sample-covariance matrix, say $\boldsymbol{S}^{(E_2)}$, of a HFM from E_2 converges to the same population matrix $\boldsymbol{\Sigma}$ as that of a SFM from G_2. However, the sampling distributions of $\boldsymbol{S}^{(E_2)}$ and $\boldsymbol{S}^{(G_2)}$ will be different. For multivariate normal data the sample-covariance matrix $\boldsymbol{S}^{(G_2)}$ follows a Wishart distribution and the usual fit statistics a central chi-square distribution (Ghosh & Sinha, 2002; Rao, 1973). If the true model is the HFM, the data do not follow a multivariate normal distribution and these distributional properties do not hold for $\boldsymbol{S}^{(E_2)}$ (Browne, 1982). However, the following simulation suggests that the distribution of the likelihood-ratio goodness-of-fit statistic (LR statistic) may not depart significantly from the central chi-square distribution if the S1FM is tested on data generated under a H1FM from E_2.

Consider four simulations. They differ in sample size, 100 or 500, and/or loading heterogeneity, $\mathrm{var}(\lambda_{ij}) = 0$ (S1FM) or $\mathrm{var}(\lambda_{ij}) = 1$ (H1FM). In each simulation, the $k = 5$ factor loadings are uncorrelated and have means $(0.9, 0.8, 0.7, 0.6, 0.5)$, the residuals and the factor each have a standard normal distribution and are uncorrelated. In each simulation 1000 samples are drawn. For each sample, the S1FM was estimated and the LR statistic computed.

Figures 1 through 4 show the Q-Q plots of the empirical distribution of the LR statistics, obtained from the 1000 samples, against the χ^2 distribution with five $(k(k+1)/2 - 2k)$ degrees of freedom. It is seen that loading heterogeneity

[1]In the discussion section we conjecture that this assumption may not always be as restrictive as it appears.

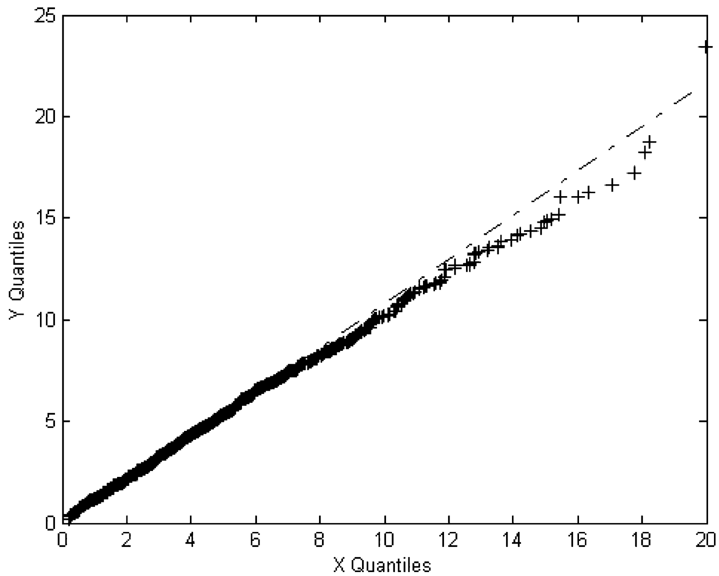

FIGURE 1 Q-Q plot expected (X) against empirical (Y), simulation: $N = 100$, S1FM data.

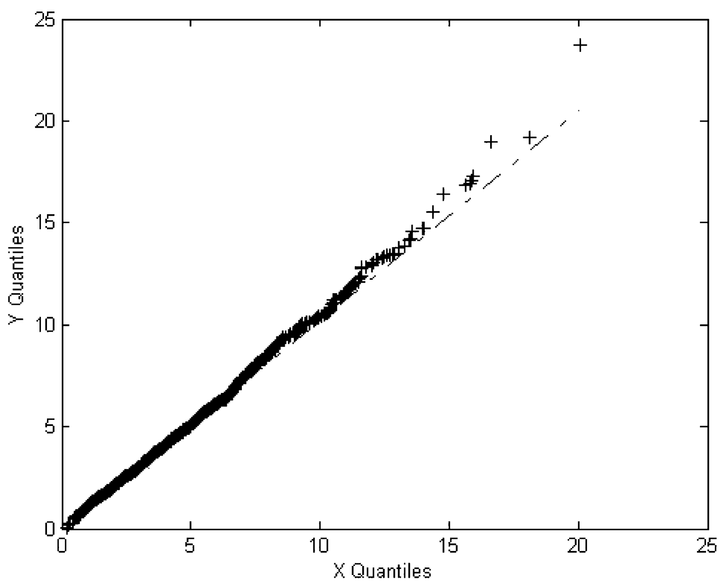

FIGURE 2 Q-Q plot expected (X) against empirical (Y), simulation: $N = 100$, H1FM data.

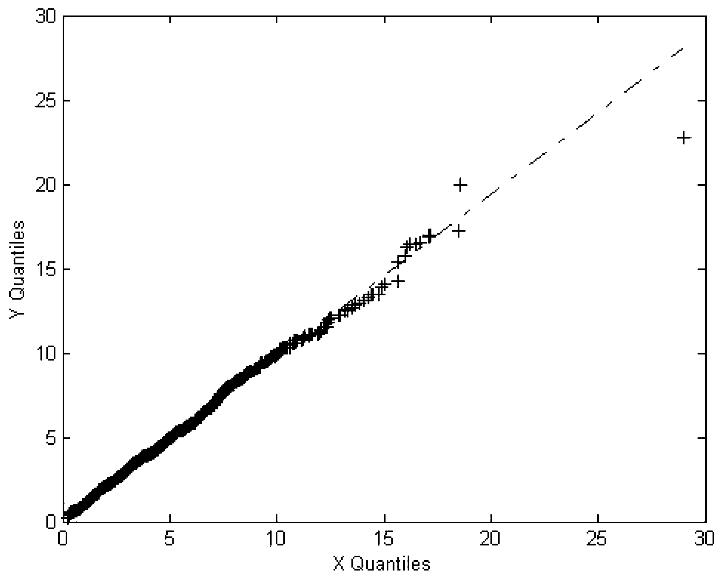

FIGURE 3 Q-Q plot expected (X) against empirical (Y), simulation: $N = 500$, S1FM data.

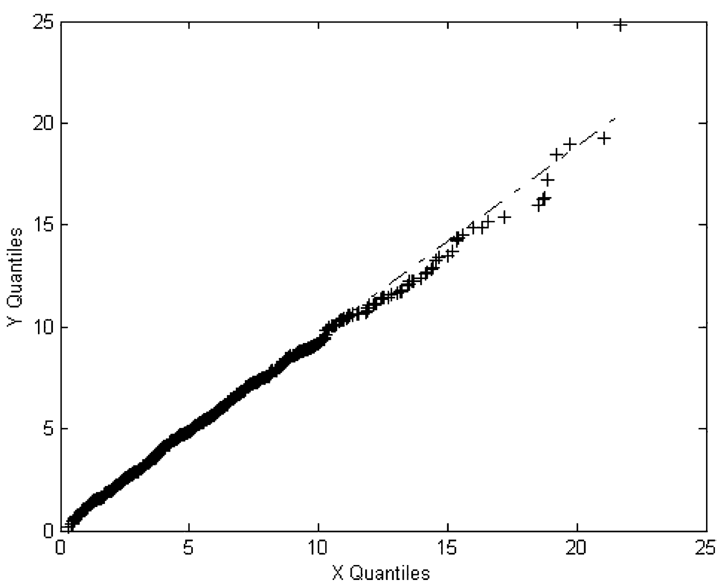

FIGURE 4 Q-Q plot expected (X) against empirical (Y), simulation: $N = 500$, H1FM data.

does not result in visible deviations of the S1FM LR statistic from χ^2_5, nor is the fit worse than for data generated under the S1FM. To test this observation statistically, two tests are performed. Table 1 gives the results of a the overall Pearson goodness-of-fit statistic (19 bins, DF = 18) for the fit of the LR statistics to the χ^2_5. This statistic tests against the vague alternative that the binned LR statistic is distributed as some arbitrary multinomial distribution. The second statistic tests whether the LR statistic is distributed as χ^2_5(= Gamma (2.5, 2)) against the alternative that it is distributed as some arbitrary Gamma (a, b), where a and b are estimated from the data. The latter test is more powerful, but makes stronger assumptions. None of the tests shows a significant deviation of the LR statistic from χ^2_5 at $\alpha = .05$. A Kolmogorov-Smirnov test gave similar results. The results indicate that the distributions of standard fit statistics under heterogeneous models do not depart significantly from what is expected under the SFM.

Furthermore, there is not a large increase in rejection rate of the SFM when the loadings are heterogeneous. For the extreme case of $\mathrm{var}(\lambda_{ij}) = 1$, the simulation results ($S = 100000$ draws) yield a percentage of correct rejections of the standard factor model of 14 and 18% for $N = 100$ and 500, respectively. Even with a sample size of $N = 5000$, the statistic rejects the SFM in only 19% of the cases. Thus, in this HFM case, one would accept the SFM in no less than 81% of the cases rather the than 95% of the cases if the data came from a SFM.

If the SFM is not rejected, one might feel justified to use its factor scores, for example for diagnostic purposes. The next section discusses the effect of

TABLE 1
Fit to χ^2_5 of LR Statistics of the S1FM for Simulated Data

	Fit statistics			
	Pearson (19 bins)		*LR Gamma(2.5, 2) against Gamma(a, b)*	
Parent Population	*Statistic*	*P (df = 18)*	*Statistic*	*P (df = 2)*
		$N = 100$		
S1FM	20.72	0.29	2.60	0.27
H1FM	16.92	0.53	0.75	0.69
		$N = 500$		
S1FM	11.29	0.88	1.17	0.65
H1FM	10.88	0.90	4.54	0.10

heterogeneity of the factor loadings on the fidelity of factor scores that are computed under the assumption that the SFM holds.

THE EFFECT OF LOADING HETEROGENEITY ON THE FIDELITY OF FACTOR SCORES

Non-rejection of the SFM with the standard LR-statistic does not guarantee the fidelity of estimated factor-scores. Figure 5 (6) gives a plot of estimated factor-scores (regression method) against the true factor-scores from the standard (heterogeneous) one-factor model, where $N = 100$ and the $\{E(\lambda_{ij})\}_j$ are as before. It is seen that in the heterogeneous model the points are somewhat more dispersed than in the standard model.

Lord and Novick (1968, p. 61) defined $\mathrm{cor}(x, \hat{x})^2$ as the *reliability* of an estimated test-score $\hat{x}$. Mellenbergh (1994, 1996) applied this measure to latent variables in factor analysis and item-response theory. If we use it to describe the fidelity of the estimated factor-scores in Figure 5 and 6, we obtain values of 0.719 and 0.487, respectively. That is, there is a marked influence of loading heterogeneity on the reliability of the factor scores.

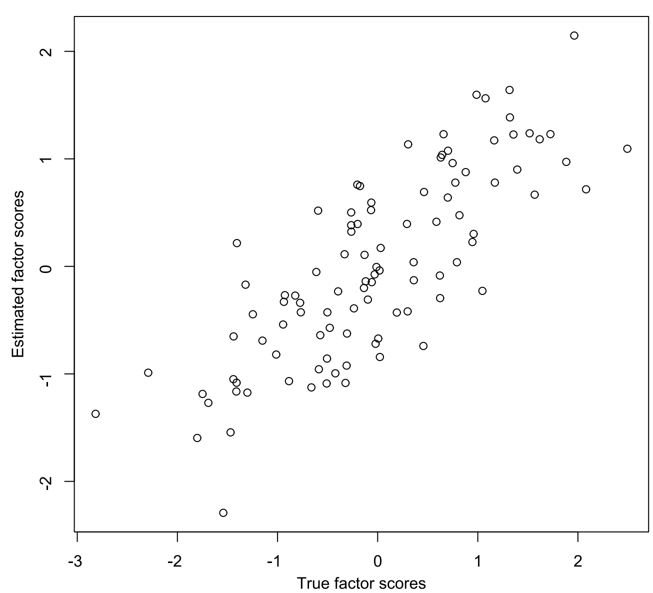

FIGURE 5 Estimated factor scores against true factor scores $\mathrm{var}(\lambda_{ij}) = 0$, simulation: $N = 100$, $n = 5$.

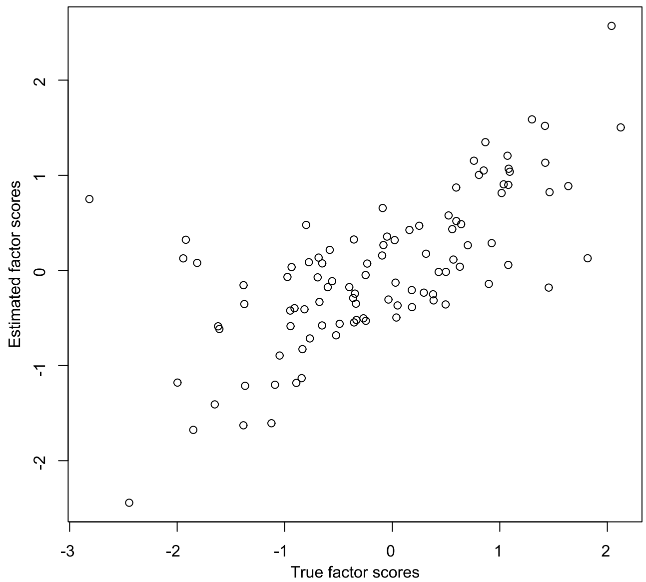

FIGURE 6 Estimated factor scores against true factor scores $\text{var}(\lambda_{ij}) = 1$, simulation: $N = 100, n = 5$.

How this factor score reliability is related to loading variance is shown in Figure 7. The mean reliabilities are plotted for simulated data sets with values $\text{var}(\lambda_{ij})$ ranging from 0 through 1. It is seen that the reliabilities depend linearly on $\text{var}(\lambda_{ij})$. Improving the estimates of λ_i by increasing the sample size does not change the linearity of the relation between factor reliability and loading variance. Comparing Figure 8 with Figure 7, it is seen that, if the sample size is taken as 500 rather than 100, the factor score reliability improves by about .02 across all loading variances. Figure 9 shows that increasing the number of items from 5 to 9 does improve the reliability but does not change the linearity of the relation.

THE EFFECT OF LOADING HETEROGENEITY ON OBSERVED SCORE DISTRIBUTIONS

The possibility to conclude that there is heterogeneity in the factor loadings depends on one's willingness to assume a specific joint distribution for η_i and $\{\epsilon_{ij}\}_j$. Because the attributes one assesses with factor analysis are often believed to emerge from a central limit effect (Billingsley, 1995, Ch. 6; Box & Tiao, 1992, Ch 4) multivariate normal distributions are usually assumed.

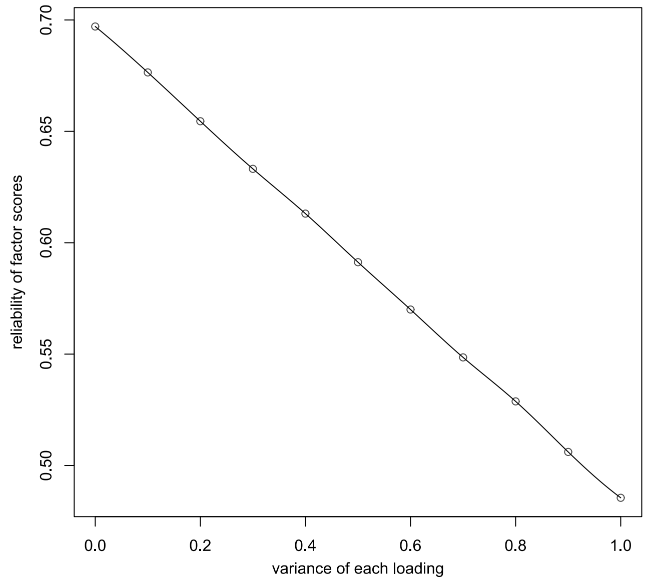

FIGURE 7 Factor score reliability against var(λ_{ij}), simulation: $N = 100$, $n = 5$.

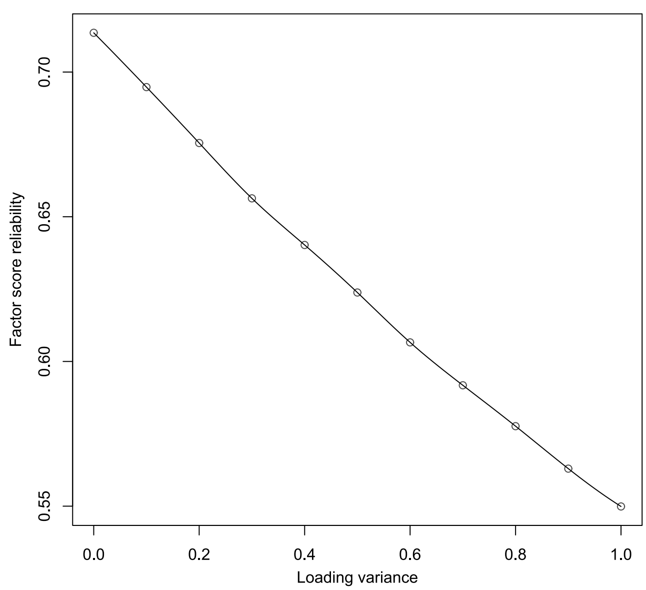

FIGURE 8 Factor score reliability against variance of loadings, simulation: $N = 500$, $n = 5$.

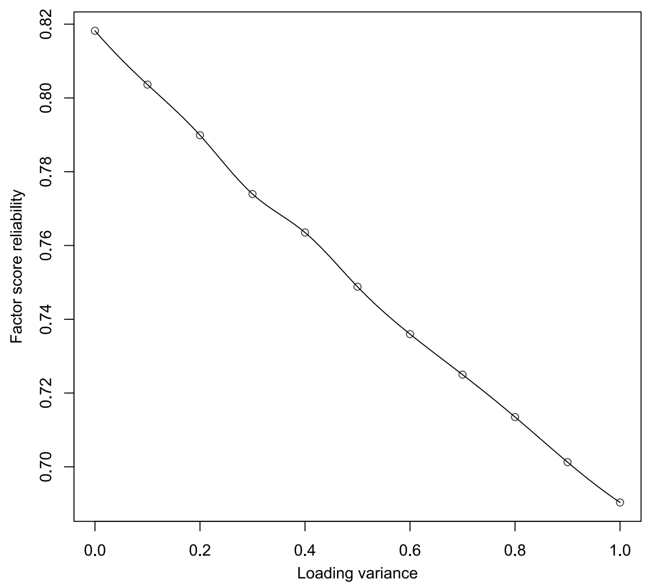

FIGURE 9 Factor score reliability against variance of loadings, simulation: $N = 500$, $n = 9$.

If loadings are statistically independent, heterogeneity expresses itself primarily in the marginal distributions $\{f(y_{ij})\}_j$. To focus exclusively on this deviation from the SFM, we limit ourselves to the study of marginal score-distributions. If λ_{ij} is fixed, $z = \lambda_{ij}\eta_i$ will have the same distribution type as η_i. Furthermore, if η_i and ϵ_{ij} are normal, y_{ij} will be too. Thus, knowledge of the distribution of η_i and ϵ_{ij} and knowing that λ_{ij} is fixed, yields knowledge of the distribution of y_{ij}. The reverse, of course, is not true. Knowledge of the distribution of y_{ij} does not uniquely identify the distribution of λ_{ij}, η_i, and ϵ_{ij}, nor which of these variables are fixed or random. So in the HFM, conclusions about the heterogeneity of factor loadings depend on assumptions about the distributions of the other random parameters in the model. If ϵ_{ij} is assumed normal, non-normality of y_{ij} implies that z is not normal. Furthermore, if η_i is assumed normal, non-normality of z, in turn, implies that λ_{ij} is not fixed. Thus, under these normality assumptions, non-normality of the marginal distribution $f(y_{ij})$ indicates loading heterogeneity.

Figure 10 shows the distribution of $z = \lambda_{ij}\eta_i$ and its normal Q-Q plot for normally distributed heterogeneous λ_{ij}. For that case, the distribution of z is studied by Craig (1936). Appendix II gives the distribution function of z for independent λ_{ij} and η_i. The distribution is rather peaked reaching its maximum if the means of λ_{ij} and η_i approach zero. Adding a normally distributed ψ_{ij} to

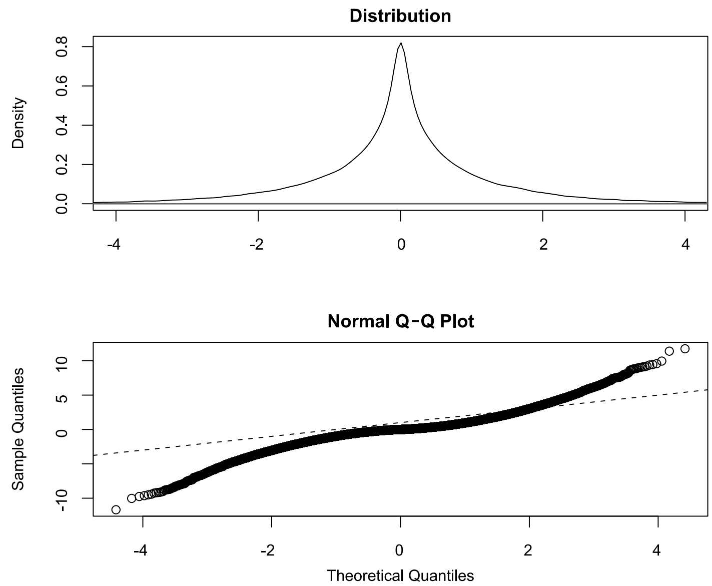

FIGURE 10 Distribution and normal Q-Q plot of $\lambda_{ij}\eta_i$ for $E(\lambda_{ij}) = 0.9$, $\text{var}(\lambda_{ij}) = \text{var}(\eta_i) = \text{var}(\psi_{ij}) = 1$, and $N = 10^5$.

obtain y_{ij} of (4), the effect is similar but less visible, see Figure 11. Appendix II gives the distribution function $f(y_{ij})$ for arbitrary means and variances of the random parameters.

A similar picture is obtained if the loadings are not normally distributed. Figure 12 shows a distribution of y_{ij} if λ_{ij} has a Chi-squared distribution on one degree of freedom. For comparability with Figure 11, the variate is re-scaled to have variance 1. It is seen from Figure 12, that the type of deviation from normality of y_{ij} is not very different from that for a normally distributed loading, given in Figure 11.

A test that is sensitive for a wide range of deviations from normality is Shapiro and Wilk's (1965) W statistic (Royston, 1982ab, 1995). W is easily performed on raw data with Slawomir Jarek's freely available function *mshapiro.test* in the R-package *mvnormtest*. In essence, the test assesses the straightness of the normal Q-Q plot of y_{ij}. For example, for a sample of $N = 100$ (500) drawn 100000 times form a heterogeneous model with $\lambda_{ij} \sim N(0.5, 1)$ the test correctly rejects our null hypothesis in 48% (98%) of the cases ($\alpha = 0.05$), whereas for a sample of ($N = 5000$) it almost always rejects the null hypothesis that λ_{ij} is fixed given that η_i and ϵ_{ij} are normal. This suggests that the power of W is quite satisfactory.

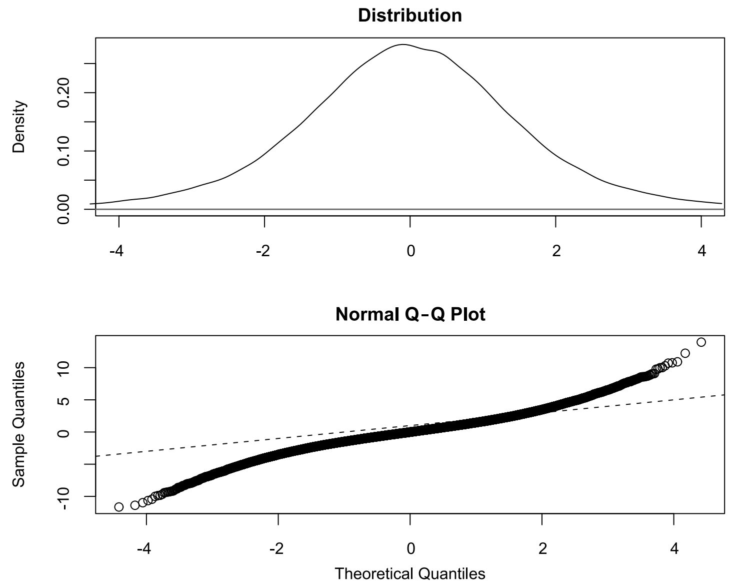

FIGURE 11 Distribution and normal Q-Q plot of y_{ij} for $\mathrm{E}(\lambda_{ij}) = 0.9$, $\mathrm{var}(\lambda_{ij}) = \mathrm{var}(\eta_i) = \mathrm{var}(\psi_{ij}) = 1$, and $N = 10^5$.

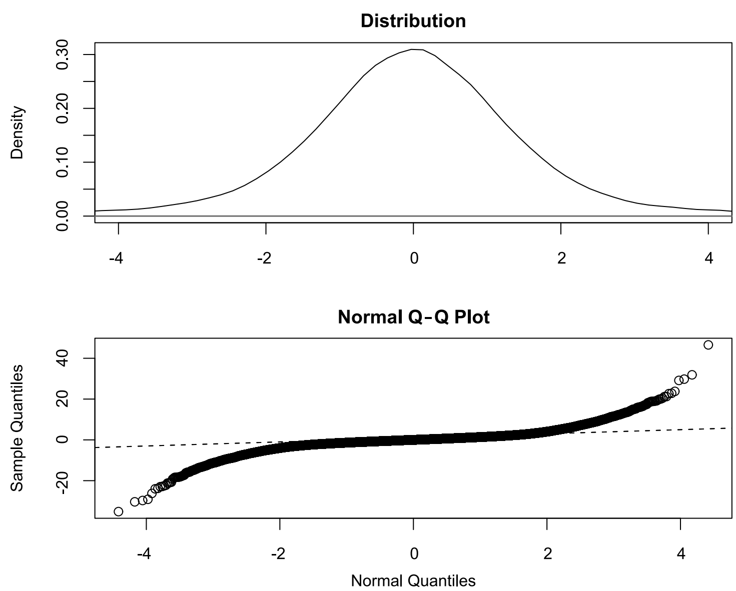

FIGURE 12 Normal Q-Q plot for y_{ij} for $\lambda_{ij} \sim \chi^2_1$ with $\mathrm{var}(\eta_i) = \mathrm{var}(\psi_{ij}) = 1$, and $N = 10^5$.

TABLE 2
Average Shapiro-Wilk Statistic W for y_{ij} over $S = 10^5$ Draws of Samples of Size of $N = 500$ from One-Factor Models with $\text{var}(\eta_i) = 1$, $\text{var}(\psi_{ij}) = 0.30$ and Normally Distributed Loadings with Various Expectations and Variances

	$E(\lambda)$									
var(λ)	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
0.1	0.9825	0.9791	0.9770	0.9771	0.9790	0.9817	0.9844	0.9871	0.9892	0.9908
0.2	0.9679	0.9651	0.9630	0.9624	0.9638	0.9668	0.9708	0.9745	0.9782	0.9814
0.3	0.9578	0.9559	0.9542	0.9531	0.9541	0.9566	0.9599	0.9643	0.9685	0.9726
0.4	0.9501	0.9489	0.9472	0.9461	0.9474	0.9492	0.9524	0.9562	0.9608	0.9650
0.5	0.9449	0.9437	0.9425	0.9418	0.9422	0.9437	0.9464	0.9498	0.9542	0.9582
0.6	0.9407	0.9396	0.9387	0.9377	0.9384	0.9393	0.9417	0.9450	0.9487	0.9529
0.7	0.9369	0.9365	0.9356	0.9347	0.9352	0.9360	0.9378	0.9408	0.9440	0.9485
0.8	0.9346	0.9339	0.9326	0.9321	0.9324	0.9331	0.9345	0.9376	0.9406	0.9443
0.9	0.9318	0.9316	0.9310	0.9301	0.9305	0.9308	0.9324	0.9348	0.9376	0.9408
1	0.9299	0.9292	0.9289	0.9284	0.9282	0.9289	0.9302	0.9323	0.9348	0.9382

Table 2 shows how W varies for normal λ_{ij} with different values of $\text{var}(\lambda_{ij})$ and $\text{E}(\lambda_{ij})$ for a factor model with the more realistic value $\text{var}(\psi_{ij}) = 0.30$ and $N = 500$. It is seen that higher values of $\text{var}(\lambda_{ij})$ lead to lower values of W, but at a different rate for different values of $\text{E}(\lambda_{ij})$. In Figure 13 the Shapiro

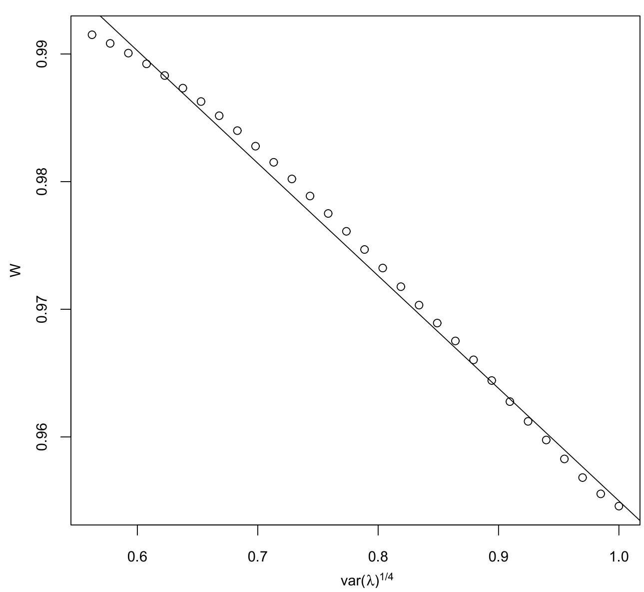

FIGURE 13 W plotted against $\text{var}(\lambda_{ij})^{\frac{1}{4}}$ for $\text{E}(\lambda_{ij}) = 0.5$ of Table 2.

Wilk statistics of Table 2 is plotted against $\text{var}(\lambda_{ij})^{\frac{1}{4}}$ together with the line of least squares fit. It is seen that for normally distributed loading variances, a near-linear relation is obtained.

Thus, Shapiro and Wilk's W statistic gives a good indication of non-normality resulting from loading heterogeneity for each observed score variable. An overall test of loading heterogeneity can be obtained by combining the W's error rates using a Bonferroni-type inequality to adjust the significance level for multiple testing (Galambos & Simonelli, 1996). It is proposed that that normality of the marginal distributions of the measurements be tested before factor scores are used to characterize the subjects.

DISCUSSION

A well-fitting factor model does not mean that the essential assumptions hold and that factor scores and their standard errors may be used for diagnostic purposes. If the factor model is heterogeneous, standard factor loadings may be interpreted as the expected value of the individuals' loadings in the population rather than a fixed loading that is invariant over individuals. It is shown that loading heterogeneity may not show up in the population covariance matrix and standard goodness of fit statistics, but does affect the fidelity of factor scores. Under the usual assumptions, loading heterogeneity results in non-normality of the measurements which can be detected with Shapiro Wilk's W test for normality. W may also be used to obtain rough estimates of loading variance using Monte Carlo methods.

If W does not reject the null hypothesis, the user may have some confidence that factor loadings are not heterogeneous. Rejection of the null hypothesis may provoke the user to generate ideas about possible causes of heterogeneity and formulate a hypothesis about the type of distribution of the loadings. For example, it may be likely that there is a central limit effect on the loading, leading to a normal distribution. In that case z would have Craig's distribution of Appendix II. If y_{ij} is a cognitive test there may be two possible solution strategies and the loadings may take one of two values depending on the solution strategy that the subject follows (Rijkes & Kelderman, 2006). In that case the loading would be binomial and z, and thus y, would become a normal mixture. Once one has a credible hypothesis about the loading distribution, a model-based approach may be pursued to test this hypothesis and estimate the relevant parameters of $f(\lambda_{ij})$.

For independent normally distributed loadings, Ansari et al. (2002) proposed a Bayesian model-based approach to assess heterogeneity of factor loadings, where replications of measurements may or may not be present. As usual, prior distributions are assumed to be independent normal with large variances to represent vague knowledge. Posterior predictive checks, comparing a heterogeneous and a standard model, are then made to test the null hypothesis that normally distributed loadings have zero variance. The intractability of the distribution of y (Appendix II) is circumvented by using Markov Chain Monte Carlo methods. These Monte Carlo methods may also be used to estimate posterior distributions of loading variances. The method may be extended to suit other loading distributions as well. It is a subject for further study to develop those methods and study their performance.

To date, it is an open question to what extend loading invariance is a problem in practice. In this paper, loading heterogeneity was introduced as a between-person effect, but model (4) may also accommodate a within-person effect. Furthermore, heterogeneous loadings may be interpreted, at the operational level, as an imperfection of the measurements, or at the theoretical level, as an indication of a phenomenon of interest, or both. An example of a within-person effect on the operational level is a random attention lapse or strategy choice during the solution of a problem (Rijkes & Kelderman, 2006). Between-person effects may result from the effects of, possibly unobserved, moderator variables on the degree to which the measurements regress on the factor. Again, this may be a flaw of the measurement instrument, such as non-uniform sex or ethnic bias, or may be interpreted substantively, such as in quantitative behavioral genetics (Martin & Eaves, 1977, Molenaar et al., 2000).

Much of the proofs and demonstrations in this paper hinge on normality assumptions. These assumptions are standard in many psychometric models. For factors and residuals, normality is an assumption that is thought to be approximately true in practice. Many populations from which we sample are quite heterogeneous and in those populations the factor distribution can arise from many different effects. According to the central limit theorem (Billingsley, 1995, Ch. 6; Box & Tiao, 1992, Ch 4), the combined effect of many different independent determinants yields a normal distribution. For a central limit effect on the factor loadings, a general biological rationale can be given that is based on the observation that behavior is dependent upon brain activity. The by far most successful class of mathematical biological models to explain the epigenetic growth of neural networks in the brain involve self-organizing reaction-diffusion models. The first model of this kind was formulated by Alan Turing (1952), followed by the classic paper of Gierer and Meinhardt (1972; cf. also Meinhardt, 1982). The reader is referred to Murray (2002) for a detailed review. Applying reaction-diffusion models of developmental instability (cf. Graham, Emlen, and Freeman, 2003) to the growth of neural networks, Molenaar, Boomsma, and Dolan (1993) showed that the resulting phenotypic variation constitutes an third source of individual differences, alongside genetic and environmental influences.

Taken together the evidence related to self-organizing epigenesis governed by reaction-diffusion models suggests that, at the micro-level, human brain architecture will be quite heterogeneous. Insofar as human behavior and information processing is dependent upon neural modules or networks, this heterogeneity can be expected to be reflected in the normality of factor loadings. It should, however, be stipulated that the same moderator variables may affect different phenotypic behaviors in a similar way causing loadings to become correlated, which is contrary to the assumptions in this paper and in Ansari et al. (2002).

Correlations between loadings may, however, be hard to identify in a HFM. In practice, the loadings may even seem to be uncorrelated because their common variance can be absorbed into the factors. To see this, suppose, for simplicity, that one only has one factor, say, a genetic factor. If the heterogeneous loadings are written as multiplicative functions of a common part ν_i ($\mathrm{var}(\nu_i) > 0$) and uncorrelated unique parts $\nu_{ij}, j = 1, \ldots, k$ ($\mathrm{cov}(\nu_{ij}, \nu_{ik}) = 0, j \neq k$), the common part can be absorbed into a new factor $\xi_i = \nu_i \eta_i$ and the corresponding H1FM becomes

$$y_{ij} = \nu_{ij}\xi_i + \psi_{ij}, \ j = 1, \ldots, n, \tag{8}$$

where the $\{\nu_{ij}\}_j$ are uncorrelated with ξ_i and with each other. However, if the effect of the moderator variable on the factor loadings is absorbed in the factor, the factor (ξ_i) would no longer be normally distributed, nor would it be a purely genetic factor. In that case, it may be better to use the approach of Purcell (2002) who introduces moderator effects in the structural model, explaining the factor's distribution, rather than that of the loadings. He assumes that the factor model is standard (1), but the factor is modeled by a variance component model that includes genetic and environmental random main effects and their interactions.

In other applications, there may be different substantive arguments to assume that the factors are not normally distributed. For example, if the population of interest consists of distinct subpopulations (e.g. diseased/healthy) for which the factor distributions differ greatly in location. Therefore, before loading heterogeneity is hypothesized, one should be able to exclude alternative violations of the essential assumptions of the factor model before performing statistical tests that are based on these assumptions.

REFERENCES

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor models: A Bayesian approach. *Psychometrika, 67*, 49–78.

Billingsley, P. (1995). *Probability and measure.* New York: Wiley.

Box, G., & Tiao, G. (1992). *Bayesian inference in statistical analysis*. John Wiley & Sons.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141), London: Cambridge University Press.

Gierer, A., & Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik, 12*, 30–39.

Graham, J. H., Emlen, J. M., & Freeman, D. C. (2003). Nonlinear dynamics and developmental instability. In M. Polak (Ed.), *Developmental instability: Causes and consequences.* (pp. 35–50). Oxford: Oxford University Press.

Craig, C. C. (1936). On the frequency function of xy. *The Annals of Mathematical Statistics, 7*, 1–15.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Galambos, J., & Simonelli, I. (1996). *Bonferroni-type inequalities with applications*. New York: Springer-Verlag.

Gardiner, C. W. (1997). *Handbook of stochastic methods for physics, chemistry and the natural sciences.* Berlin: Springer.

Ghosh, M., & Sinha, B. K. (2002). A simple derivation of the Wishart distribution. *The American Statistician, 56*, 100–101.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 109–133.

Kendall, M. G., Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics: Volume 1, distribution theory.* New York: Oxford University Press.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* London: Butterworth.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity, 38*, 79–95.

McCullagh, P. (1987). *Tensor methods in statistics.* London: Chapman and Hall.

Meinhardt, H. (1982). *Models of biological pattern formation*. London: Academic Press.

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29*, 223–236.

Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods, 1*, 293–299.

Molenaar, P. C. M., Boomsma, D. I., & Dolan, C. V. (1993). A third source of developmental differences. *Behavior Genetics, 23*, 519–524.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2000). The relationship between the structure of inter-individual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & Lindenberger (Eds.), *Understanding human development*. New York: Kluwer Academic Publishers.

Murray, J. D. (2002). *Mathematical biology II.* 3rd edition. New York: Springer Verlag.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*, 554–571.

Rao, C. R. (1973). *Linear statistical inference and its applications*. London: Wiley

Rijkes, C. P. M., & Kelderman, H. (2006). Latent response Rasch models for strategy shifts in problem solving processes. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications.* (pp. 313–332). New York: Springer Verlag.

Royston, P. (1982a). Algorithm AS 181: The W test for normality. *Applied Statistics, 31*, 176–180.

Royston, P. (1982b). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics, 31*, 115–124.

Royston, P. (1995). A remark on algorithm AS 181: The W test for normality. *Applied Statistics, 44*, 547–551.

Shapiro, S. S., & Wilk, M. B. (1965). An analyis of test for normality (complete samples). *Biometrika, 52*, 591–611.

Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology, 5*, 201–292.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society London, Series B, 237*, 37–72.

APPENDIX I

Proof of Theorem 1. Use the cumulant (κ) expansion of expectations (McCullagh, 1987, p. 29) and the fact that for the multivariate normal distribution cumulants of order 1 and 2 satisfy $\kappa(x_i) = \mathrm{E}(x_i)$ and $\kappa(x_i, y_i) = \mathrm{cov}(x_i, y_i)$, and all cumulants of higher order than 2 vanish (Gardiner, 1997, p. 34; Kendall, Stuart, & Ord, 1994, p. 93). For brevity denote $\beta_{ij} \equiv \mathrm{D}(\lambda_{ij})$ and use $\mathrm{E}(\beta_{ij}) = 0$, and, if $\mathrm{E}(x_i) = 0$, use $\mathrm{E}(\beta_{ij} x_i) = \mathrm{cov}(\beta_{ij}, x_i)$.

From Lemma 1, the properties of cumulants, and assumptions (4), (5), and (6), one obtains

$$
\begin{aligned}
\mathrm{cov}(\eta_i, \epsilon_{ij}) &= \mathrm{E}\{\eta_i \mathrm{D}(\beta_{ij}\eta_i)\} + \mathrm{E}(\eta_i \psi_{ij}) - \mathrm{E}(\eta_i)\mathrm{E}(\psi_{ij}) \\
&= \mathrm{E}(\eta_{ij}^2 \beta_{ij}) - \mathrm{E}\{\eta_i \mathrm{E}(\eta_i \beta_{ij})\} + 0 \\
&= \kappa(\eta_i, \eta_i, \beta_{ij}) + 2\kappa(\eta_i)\kappa(\eta_i, \beta_{ij}) + \kappa(\eta_i, \eta_i)\kappa(\beta_{ij}) \\
&\quad + \kappa(\eta_i)\kappa(\eta_i)\kappa(\beta_{ij}) \\
&= 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{cov}(\epsilon_{ij}, \epsilon_{ik}) &= \mathrm{E}[\{\mathrm{D}(\beta_{ij}\eta_i) + \psi_{ij}\}\{\mathrm{D}(\beta_{ik}\eta_i) + \psi_{ik}\}] \\
&\quad - \mathrm{E}\{\mathrm{D}(\beta_{ij}\eta_i) + \psi_{ij}\}\mathrm{E}\{\mathrm{D}(\beta_{ik}\eta_i) + \psi_{ik}\} \\
&= \mathrm{E}\{\mathrm{D}(\psi_{ij})\mathrm{D}(\beta_{ik}\eta_i)\} + \mathrm{E}\{\mathrm{D}(\psi_{ik})\mathrm{D}(\beta_{ij}\eta_i)\} \\
&\quad + \mathrm{E}\{\mathrm{D}(\beta_{ij}\eta_i)\mathrm{D}(\beta_{ik}\eta_i)\} \\
&\quad - \mathrm{E}\{\mathrm{D}(\beta_{ij}\eta_i)\}\mathrm{E}\{\mathrm{D}(\beta_{ik}\eta_i)\}
\end{aligned}
$$

where,

$$\begin{aligned}
\mathrm{E}\{\mathrm{D}(\psi_{ij})\mathrm{D}(\beta_{ik}\eta_i)\} &= \mathrm{E}(\psi_{ij}\beta_{ik}\eta_i) - \mathrm{E}(\psi_{ij})\mathrm{E}(\beta_{ik}\eta_i) \\
&= \kappa(\psi_{ij}, \beta_{ik}, \eta_i) + \kappa(\psi_{ij})\kappa(\beta_{ik}, \eta_i) \\
&\quad + \kappa(\beta_{ik})\kappa(\psi_{ij}, \eta_i) \\
&\quad + \kappa(\eta_i)\kappa(\beta_{ik}, \psi_{ij}) + \kappa(\psi_{ij})\kappa(\beta_{ik})\kappa(\eta_i) \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\{\mathrm{D}(\beta_{ij}\eta_i)\mathrm{D}(\beta_{ik}\eta_i)\} &= \mathrm{E}(\beta_{ij}\beta_{ik}\eta_i^2) - \mathrm{E}(\beta_{ij}\eta_i)\mathrm{E}(\beta_{ik}\eta_i) \\
&= \kappa(\beta_{ij}, \eta_i, \beta_{ik}, \eta_i) + \kappa(\beta_{ij})\kappa(\eta_i, \beta_{ik}, \eta_i) \\
&\quad + 2\kappa(\eta_i)\kappa(\beta_{ij}, \beta_{ik}, \eta_i) + \kappa(\beta_{ik})\kappa(\eta_i, \beta_{ij}, \eta_i) \\
&\quad + 2\kappa(\beta_{ij}, \eta_i)\kappa(\beta_{ik}, \eta_i) + \kappa(\beta_{ij}\beta_{ik})\kappa(\eta_i, \eta_i) \\
&\quad + \kappa(\beta_{ij})\kappa(\eta_i)\kappa(\beta_{ik})\kappa(\eta_i) \\
&\quad - \mathrm{cov}(\beta_{ij}, \eta_i)\mathrm{cov}(\beta_{ik}, \eta_i) \\
&= \mathrm{cov}(\beta_{ij}, \beta_{ik})\mathrm{cov}(\eta_i, \eta_i) + \mathrm{cov}(\beta_{ij}, \eta_i)\mathrm{cov}(\beta_{ik}, \eta_i) \\
&= \mathrm{cov}(\beta_{ij}, \beta_{ik})\mathrm{var}(\eta_i) + \mathrm{cov}(\beta_{ij}, \eta_i)\mathrm{cov}(\beta_{ik}, \eta_i) \\
&= \mathrm{cov}(\beta_{ij}, \beta_{ik}) + \mathrm{cov}(\beta_{ij}, \eta_i)\mathrm{cov}(\beta_{ik}, \eta_i),
\end{aligned}$$

$$\mathrm{E}\{\mathrm{D}(\beta_{ij}\eta_i)\}\mathrm{E}\{\mathrm{D}(\beta_{ik}\eta_i)\} = \mathrm{cov}(\beta_{ij}\eta_i)\mathrm{cov}(\beta_{ik}\eta_i),$$

so that

$$\mathrm{cov}(\epsilon_{ij}, \epsilon_{ik}) = \mathrm{cov}(\lambda_{ij}, \lambda_{ik}) = 0.$$

Consequently, the assumptions (1) and (2) of the S1FMs are satisfied by the H1FMs from E_2 and yield the covariance structure

$$\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{\lambda}_i)\Phi\mathrm{E}(\boldsymbol{\lambda}_i)^T + \boldsymbol{\Omega}_{\boldsymbol{\epsilon\epsilon}} = \boldsymbol{\lambda}\Phi\boldsymbol{\lambda}^T + \boldsymbol{\Omega}_{\boldsymbol{\epsilon\epsilon}}, \tag{9}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\{y_{ij}\}_j$, $\boldsymbol{\Omega}_{\boldsymbol{\epsilon\epsilon}}$ is the diagonal matrix of variances $\{\mathrm{var}(\epsilon_{ij})\}_j$, Φ is the variance of η_i, $\boldsymbol{\lambda}_i = \{\lambda_{ij}\}_j$, and $\boldsymbol{\lambda} = \{\lambda_j\}_j$.

APPENDIX II

Omitting indices, write Equation 4 as $y = z + \psi$, with $z = \lambda\eta$, where λ, η and ψ are independent normal variables with means μ_λ, μ_η and μ_ψ and variances σ_λ^2, σ_η^2 and σ_ψ^2 respectively.

The distribution of z is obtained as follows. First note that

$$f(z) = \int_{-\infty}^{\infty} f(z,\eta)d\eta = \int_{-\infty}^{\infty} f(\eta)f(z|\eta)d\eta. \tag{10}$$

One has

$$f(\eta) = \frac{1}{\sqrt{2\pi}\sigma_\eta} e^{-\frac{1}{2\sigma_\eta^2}(\eta-\mu_\eta)^2} \tag{11}$$

and, from standard theory of conditional normal distributions,

$$f(z|\eta) = \frac{1}{\sqrt{2\pi} \mid \eta \mid \sigma_\eta} e^{-\frac{1}{2\eta^2\sigma_\lambda^2}(z-\eta\mu_\lambda)^2}, \qquad (z,\eta) \neq (0,0). \tag{12}$$

To see this, note that in $f(\lambda\eta|\eta)$, λ is the only random variable. In fact, in this distribution, η is a fixed multiplicative scale factor that transforms λ onto z. Consequently, $f(z|\eta) = f(\lambda\eta|\eta)$ is normal with mean $\eta\mu_\lambda$ and variance $\eta^2\sigma_\lambda^2$. Note that there is an singularity in $f(z,\eta)$ at $(z,\eta) = (0,0)$.

Substituting (11) and (12) in (10) yields, after simplification,

$$f(z) = \int_{\Re-\{0\}} \frac{1}{2\pi \mid \eta \mid \sigma_1\sigma_\eta} e^{-\frac{1}{2\sigma_\eta^2}(\eta-\mu_\eta)^2-\frac{1}{2\sigma_\lambda^2}(\frac{z}{\eta}-\mu_\lambda)^2} d\eta. \tag{13}$$

To obtain the distribution function of y, note that, because both λ and η are independent of ψ, $z = \lambda\eta$ is independent of ψ, so

$$f(z,\psi) = f(z)f(\psi). \tag{14}$$

If $A(y)$ denotes the set of values of (z,ψ) for which $z + \psi = y$ we have distribution function

$$f(y) = \int_{A(y)}\int_{\Re-\{0\}} \frac{1}{(2\pi)^{\frac{3}{2}} \mid \eta \mid \sigma_\lambda\sigma_\eta\sigma_\psi} \times e^{-\frac{1}{2\sigma_\eta^2}(\eta-\mu_\eta)^2-\frac{1}{2\sigma_\lambda^2}(\frac{z}{\eta}-\mu_\lambda)^2-\frac{1}{2\sigma_\psi^2}(\psi-\mu_\psi)^2} d(z,\psi)d\eta, \tag{15}$$

which is quite intractable and is best studied with computational means.